



Universidade de Brasília
IE - Departamento de Estatística
Estágio Supervisionado II

Detecção de Clusters

Uma análise de sazonalidade de surtos de dengue nos
municípios do Brasil de 2007 a 2011

Rafael Lewkowicz

Relatório Final

Orientador: Prof. Dr. André Luiz Fernandes Cançado

Brasília

Janeiro de 2013

Rafael Lewkowicz

Detecção de Clusters

Uma análise de sazonalidade de surtos de dengue nos
municípios do Brasil de 2007 a 2011

Monografia apresentada à Banca designada
pelo departamento de estatística da Univer-
sidade de Brasília, como requisito final para
a obtenção do grau de BACHAREL em Es-
tatística.

Orientador: Prof. Dr. André Luiz Fernandes Cançado

Brasília

2013

Resumo

O trabalho aqui proposto objetiva a verificação e validação da hipótese de sazonalidade na ocorrência de casos de dengue no território brasileiro, utilizando a estatística de varredura (Estatística Scan) proposta por Kulldorff (1997) para detecção e identificação de clusters temporais. Os resultados apontam que existe, na maioria das vezes, uma concentração dos casos no período de janeiro a maio destes anos. Uma vez que a distribuição dessa estatística sob H_0 é desconhecida, a significância da solução é avaliada tomando-se uma amostra suficientemente grande dessa distribuição, por meio de simulações de Monte Carlo.

Palavras-chave: Clusters, Conglomerados, Espacial, Temporal, Espaço-Temporal, sazonalidade, Dengue, Kulldorff, Verossimilhança, Monte, Carlo, Estatística Scan, Satscan.

Abstract

The work proposed here aims to verify and validate the hypothesis of seasonality in the occurrence of dengue cases in the Brazilian territory, using the Scan Statistics proposed by Kulldorff (1997) for detection and identification of temporal clusters. The results indicate that there is, in most cases, a concentration of cases from January to May of these years. Since the distribution of this statistic under H_0 is unknown, the significance of the solutions is evaluated by taking a sufficiently large sample of this distribution through Monte Carlo simulations.

Keywords: Clusters, Cluster, Spatial, Temporal, Spatio-Temporal, seasonality, Dengue, Kulldorff, Likelihood, Monte Carlo, Statistics Scan, Satscan.

Sumário

1	Introdução	5
1.1	Motivação	7
1.2	Formulação do problema e Objetivos	8
1.3	Metodologia utilizada	8
2	Estatística Scan	9
2.1	Estatística Scan Espacial	9
2.1.1	O teste de hipóteses	9
2.1.2	Estatística do teste de razão de verossimilhança	10
2.2	Estatística Scan Temporal	13
2.3	Estatística Scan Espaço-temporal	13
2.4	Candidatos a conglomerados	14
2.4.1	Caso Espacial	14
2.4.2	Caso Temporal	14
2.4.3	Caso Espaço-Temporal	15
2.5	Simulação de Monte Carlo	15
3	O Banco de dados	17
3.1	Problema em questão	17
3.2	Os dados e o DataSUS	17
4	Discussão dos Resultados	20
4.1	Análise Puramente Temporal	20
4.1.1	Apresentação do output do software	20
4.1.2	Análise e Discussão	22
4.2	Análise Espaço-Temporal	25
5	Considerações finais e trabalhos futuros	31
5.1	Considerações finais	31
5.2	Trabalhos futuros	31
A	Output do Software	33
A.1	Ano de 2007	33
A.2	Ano de 2008	35
A.3	Ano de 2009	37
A.4	Ano de 2010	39
A.5	Ano de 2011	41
A.6	Intervalo completo 2007 - 2011 (Scan Temporal)	43
A.7	Intervalo completo 2007 - 2011 (Scan Espacial)	45

Capítulo 1

Introdução

Desde a década de 80, o Brasil vem enfrentando grandes dificuldades para controlar as infecções produzidas pelo vírus da dengue que passou a circular com intensidade surpreendente, em grandes e pequenos centros urbanos, produzindo epidemias sucessivas de elevada magnitude no território nacional.

Em termos de saúde pública, a dengue vem sendo um dos maiores problemas que nosso governo vem encontrando neste século. As epidemias relâmpago causadas pelo vírus mostraram-se sair do controle por volta de 2002 até que, em 2008, a doença volta com força assustadora. No Brasil, está-se diante de uma epidemia de grandes proporções, suscetível a estender-se por anos apesar dos múltiplos esforços que se faz para contê-la, e que pode agravar-se devido a um aumento de custos para mantê-la sob controle gerando crises, e por sua vez desequilíbrio no sistema de saúde.

Este trabalho trata do estudo dos casos de dengue nos municípios brasileiros no período de 2007 a 2011 em busca de clusters temporais para a verificação da hipótese de sazonalidade na variação da incidência da doença, ou seja, buscamos saber se a ocorrência da doença se dá de modo meramente aleatório ou se segue algum padrão que gere um aglomerado anormal de casos em algum período de tempo do ano. Segundo Hair (1998):

A análise de cluster, também conhecida como análise de conglomerados, é um conjunto de técnicas estatísticas com o objetivo de agrupar objetos segundo suas semelhanças, formando grupos ou conglomerados homogêneos. Os objetos em cada conglomerado tendem a ser semelhantes entre si, porém diferentes dos demais objetos dos outros conglomerados. Os conglomerados devem ser homogêneos internamente (dentro de cada conglomerado), em conjunto com uma grande heterogeneidade externa (entre conglomerados). Portanto, se a aglomeração for bem sucedida, quando representados graficamente, os objetos dentro dos conglomerados estarão muito próximos, e os conglomerados distintos estarão afastados.

Os clusters aqui estudados, no entanto, têm outra natureza. Trata-se de conglomerados temporais e espaço-temporais, que constituem-se de conjuntos de regiões geográficas adjacentes em janelas de tempo específicas. Buscamos conglomerados que concentrem uma quantidade de casos do fenômeno em estudo maior do que a esperada, constituindo, portanto, uma anomalia.

Quando se tratando de doenças contagiosas no geral, assim como a dengue, é necessário um investimento governamental em medidas preventivas como saneamento básico, busca por focos, diagnóstico e tratamento de pacientes infectados, assim como a orientação à população a respeito de formas de prevenção e medidas de higiene. Logo, sendo até mesmo fatal em alguns casos, como no caso da febre hemorrágica, esta doença exige grande cautela e permanente fortalecimento da uma rede de vigilância epidemiológica do governo para uma boa detecção e a adoção de medidas eficazes de prevenção para controle da doença.

Para maior conhecimento do comportamento desta doença, visamos, assim, neste trabalho aplicar os procedimentos relativos à estatística Scan proposta por Kulldorff (1998) para detectar padrões da doença ao longo dos anos, buscando verificar sua sazonalidade. Esta ferramenta de Kulldorff consiste em varrer de modo exaustivo nossos intervalos de tempo, de 2007 a 2011, com janelas variáveis em busca de clusters com alta incidência. O cluster de alta incidência é o período que possui incidência significativa¹ se comparado à incidência esperada neste período.

Será realizado um teste de hipóteses para verificar se o cluster com incidência alta é, ou não, estatisticamente significativo utilizando simulações de Monte Carlo para auxiliar na decisão dos testes.

O método para obter aproximações numéricas de funções complexas que se baseia, neste trabalho, na geração de observações de alguma distribuição de probabilidades e o uso da amostra obtida para aproximar a função de interesse, é conhecido como Método de Monte Carlo. Sua origem data de 1949, com a publicação do artigo “The Monte Carlo Method” (Metropolis, 1949). A denominação do método provém do nome da cidade do principado de Mônaco. O princípio do Método de Monte Carlo já era conhecido antes da publicação do artigo de Metropolis: era utilizado, por exemplo, no tratamento de dados de amostras aleatórias em estatística. Mas a sua ampla aplicação não era viável antes do aparecimento dos computadores eletrônicos (Fishman, 1995). No projeto de construção da bomba atômica², Ulam, von Neumann e Fermi consideraram a possibilidade de utilizar o método, que envolvia a simulação direta de problemas de natureza probabilística relacionados com o coeficiente de difusão do neutron em certos materiais.

Para a análise proposta neste trabalho será utilizado um banco de dados obtido a partir do domínio público DataSus (www.datasus.gov.br) que contém os casos de incidência da doença no período de 2007 a 2011. O banco de dados consiste em uma série temporal de casos da doença. De modo sucinto, uma série temporal é qualquer conjunto de observações ordenadas no tempo, ou seja, consiste em um conjunto de observações geradas sequencialmente no tempo. Uma série temporal pode ser discreta ou contínua, dependendo se o conjunto de observações é discreto ou contínuo, respectivamente. É possível, todavia, obter uma série temporal discreta por meio da amostragem de uma série contínua em intervalos de tempo iguais. As séries que serão estudadas neste trabalho são discretas tendo em vista que as observações disponíveis são dadas em tempo discreto e em intervalos equidistantes no tempo.

O algoritmo para a análise já está implementado no software SatScan (2012), e o objetivo será de verificar a existência ou não de um, ou mais, conglomerados temporais de casos de dengue ao longo dos anos na região.

¹a nível de significância 5%.

²Conhecido também como Projeto Manhattan.

1.1 Motivação

O vírus da dengue é transmitido pela picada da fêmea do *Aedes aegypti*, o famoso mosquito diurno que se multiplica em depósitos de água parada. Constantemente vemos a dengue sendo noticiada em muitos telejornais assim como em foco frequentemente em várias revistas nacionais e internacionais, nas quais sempre é ressaltada a importância de estudar focos da doença em tempo real. O enfoque da mídia na dengue e seu alastramento pelo território nacional vêm crescendo proporcionalmente às suas epidemias.

Entre 1990 e 2010 produziram-se várias epidemias, principalmente nos grandes centros urbanos do sudeste e o nordeste do Brasil, onde se concentraram a maioria dos casos notificados. As regiões centro-oeste e norte foram afetadas posteriormente por epidemias de dengue a partir da segunda metade da década de 1990. Este crescimento absurdo está ilustrado na figura 1.1 que aponta os municípios infestados pela epidemia em 1995 e seu crescimento até 2010.

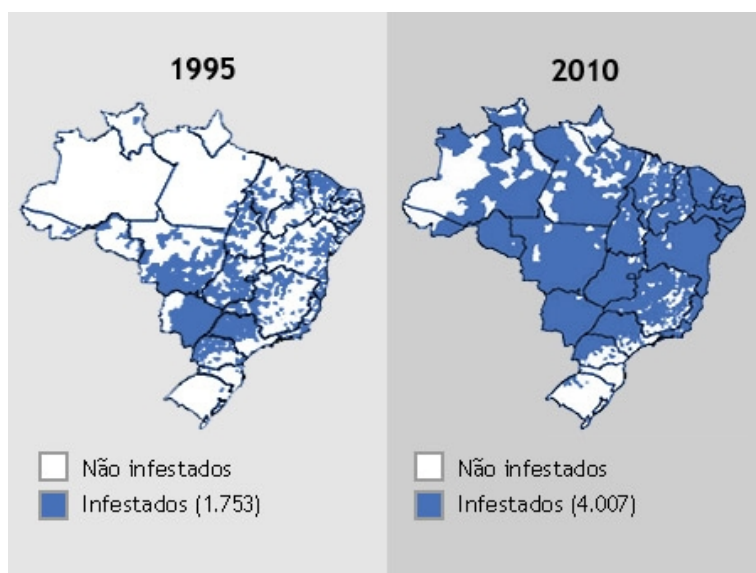


Figura 1.1: O crescimento espantoso das epidemias. Fonte: DataSus

A gravidade da situação atual da dengue no Brasil é que reveste características de epidemia; que pode ser definida como um aumento do número de casos na mesma região. A taxa de impacto da dengue em nosso país, assim como seus custos, é a mais alta que houve no mundo durante o último século.

É importante que o setor da saúde desempenhe um papel protagonista na direção das campanhas de luta antivetorial uma vez que a quase totalidade da população é de nível tanto cultural quando educacional baixo e dificilmente se preocuparia com o controle de tal epidemia quando sua primordial necessidade é preocupar-se com sua própria subsistência. A vigilância epidemiológica precisa de uma base científica forte para a detecção precoce dos surtos e capacitação do pessoal médico para conseguir um diagnóstico oportuno e certo que evite óbitos.

1.2 Formulação do problema e Objetivos

O problema abordado neste trabalho consiste em verificar se há algum padrão temporal na ocorrência de casos de dengue no território brasileiro, ou se esses casos ocorrem de forma puramente aleatória ao longo do tempo. O objetivo é verificar, portanto, se os surtos de dengue ocorrem, por exemplo, em determinada época de cada ano, ou se há maior concentração de casos em algum intervalo de tempo específico.

Detectada a presença de alguma concentração anormal, queremos identificar a janela temporal que a contenha, bem como verificar sua significância estatística, por meio de testes de hipóteses.

1.3 Metodologia utilizada

Pretende-se utilizar a estatística de varredura proposta por Kulldorff (1997) para detecção e identificação de conglomerados temporais. Uma vez que a distribuição dessa estatística sob H_0 é desconhecida, a significância das eventuais soluções será avaliada tomando-se uma amostra suficientemente grande dessa distribuição, por meio de simulações de Monte Carlo.

Capítulo 2

Estatística Scan

Aqui será detalhado o procedimento scan utilizado neste trabalho. A estatística scan (Kulldorff, 1997) é amplamente utilizada em problemas de detecção e inferência de clusters em situações onde as ocorrências no mapa são supostas aleatórias. Dentre inúmeras aplicações que este método teve, podemos citar de modo ilustrativo os trabalhos de Payão Pellegrini (2002), Chaput (2002) e Osei (2008), todos na área da epidemiologia.

A importância deste estudo é propor um algoritmo que estude, e gere resultados fiéis, de situações reais. O processo iterativo varre o mapa em busca de clusters, verificando as respectivas verossimilhanças em busca de candidatos a clusters e depois verificando suas significâncias através de uma simulação de Monte Carlo.

2.1 Estatística Scan Espacial

Com o mapa dividido em várias regiões, uma zona z do mapa é definida como um conjunto conexo de regiões. Esta zona em estudo possui uma população n_z e um número de casos ocorridos C_z com probabilidade p para que um indivíduo nesta zona seja uma ocorrência. Assim, a média de casos em z é definida por $\mu_z = pn_z$, ou seja, a média é proporcional à sua população.

Utilizaremos o modelo Poisson para descrever a distribuição de casos entre as regiões do mapa. Consideramos que o número de casos C_z é uma variável aleatória com distribuição Poisson cujo parâmetro λ_z , que representa o número esperado de casos, é tal que $\lambda_z = pn_z$. Assim:

$$C_z \sim \text{Poisson}(\lambda_z = pn_z)$$

2.1.1 O teste de hipóteses

Utilizamos um teste de hipóteses com o fim de verificar se o número de casos de dengue na área selecionada é significativo. Assim, supomos que os casos seguem uma distribuição Poisson (distribuição binomial poderia ser utilizada, porém ambas se aproximam), uma vez que ela se adequa ao ser utilizada para estudo de incidências em um determinado período de tempo. Assim o estudo do número de casos C_i em uma região r_i assume a seguinte distribuição de probabilidades:

$$f_i(c) = \frac{e^{-\lambda_i} \lambda_i^c}{c!}, c \geq 0$$

Portanto, a função de distribuição de casos C_z em uma zona z determinada se dará por:

$$f_z(c) = \frac{e^{-\mu_z} \mu_z^c}{c_z!}, c \geq 0 \quad (2.1)$$

Adotando-se N como a população total no mapa e C como o número total de casos observados, temos $n_{\bar{z}}$ e $c_{\bar{z}}$ como a população e o número de casos fora da zona z . Assim:

$$\begin{aligned} N_{\bar{z}} &= N - n_z \\ C_{\bar{z}} &= C - c_z \end{aligned}$$

Agora, partindo da hipótese de não existência de conglomerados no mapa (ou seja, total aleatoriedade da incidência de casos de dengue no mapa), definimos H_0 e H_1 . Assim é utilizado o teste de hipóteses para a existência, ou não, de conglomerados nas áreas do mapa:

$$\begin{aligned} H_0 &: p = q \\ H_1 &: p > q \end{aligned} \quad (2.2)$$

onde p é a probabilidade de que um indivíduo em z venha a ser um caso, e q é a probabilidade de que um indivíduo fora de z venha a ser um caso. Assim, pode-se concluir de modo lógico, que a zona z será considerada um conglomerado caso o teste rejeite a hipótese nula H_0 .

A estatística do teste que será utilizada neste trabalho será a razão de verossimilhança, e será detalhada na próxima seção.

2.1.2 Estatística do teste de razão de verossimilhança

Como visto em 2.2, para H_0 temos que:

$$\mu_z = pn_z \rightarrow \mu_{\bar{z}} = p(N - n_z)$$

Assim, expressamos a verossimilhança por:

$$\begin{aligned} L_0(z; p) &= \frac{\mu_z^{c_z} e^{-\mu_z}}{c_z!} \frac{\mu_{\bar{z}}^{c_{\bar{z}}} e^{-\mu_{\bar{z}}}}{c_{\bar{z}}!} \\ &= \frac{(pn_z)^{c_z} e^{-pn_z}}{c_z!} \frac{(p(N - n_z))^{(C - c_z)} e^{-p(N - n_z)}}{(C - c_z)!} \end{aligned}$$

Temos $l_0(z; p)$ como sendo o logaritmo da verossimilhança $L_0(z; p)$:

$$l_0(z; p) = c_z[\log p + \log n_z] - pn_z - \log c_z! + (C - c_z)[\log p + \log(N - n_z)] - p(N - n_z) - \log[(C - c_z)!]$$

Aplicamos a derivada de $l_0(z; p)$ uma vez que queremos encontrar o ponto que maximiza a função de verossimilhança. Logo:

$$\begin{aligned} \frac{\partial l_0(z; p)}{\partial p} &= \frac{c_z}{p} - n_z + \frac{C - c_z}{p} - (N - n_z) = 0 \\ &= \frac{C}{p} - N = 0 \\ &= p = \frac{C}{N} \end{aligned}$$

Agora substitui-se $p = \frac{C}{N}$ em $L_0(z; p)$:

$$L_0(z) = \frac{(\frac{C}{N}n_z)^{c_z} e^{-(\frac{C}{N}n_z)}}{c_z!} \frac{(\frac{C}{N}(N - n_z))^{C-c_z} e^{-(\frac{C}{N}(N - n_z))}}{(C - c_z)!}$$

Considere $\mu_z = \frac{C}{N}n_z$:

$$L_0(z) = \frac{\mu_z^{c_z} e^{-\mu_z}}{c_z!} \frac{(C - \mu_z)^{C-c_z} e^{-(C-\mu_z)}}{(C - c_z)!} = \frac{\mu_z^{c_z} (C - \mu_z)^{C-c_z} e^{-C}}{c_z! (C - c_z)!}$$

De acordo com a nossa hipótese alternativa ($H_1 : p > q$), temos que a probabilidade de que um indivíduo venha a ser uma ocorrência é maior dentro que fora da zona z :

$$\mu_z = pn_z \rightarrow \mu_{\bar{z}} = q(N - n_z), p > q$$

Calculamos, agora, a função de verossimilhança sob a hipótese H_1 (2.2):

$$L(z; p; q) = \frac{(pn_z)^{c_z} e^{-pn_z}}{c_z!} \frac{[q(N - n_z)]^{(C-c_z)} e^{-q(N-n_z)}}{(C - c_z)!}$$

Sendo $l(z; p; q) = \log L(z; p; q)$,

$$l(z; p; q) = c_z[\log p + \log n_z] - pn_z - \log c_z! + (C - c_z)[\log q + \log(N - n_z)] - q(N - n_z) - \log[(C - c_z)!]$$

Novamente, maximizando $l(z; p; q)$, temos:

$$\begin{aligned}\frac{\partial l(z; p; q)}{\partial p} &= \frac{c_z}{p} - n_z = 0 \\ &= p = \frac{c_z}{n_z} \\ \frac{\partial l(z; p; q)}{\partial q} &= \frac{C - c_z}{q} - (N - n_z) = 0 \\ &= q = \frac{(C - c_z)}{(N - n_z)}\end{aligned}$$

Agora basta aplicarmos $p = \frac{c_z}{n_z}$ e $q = \frac{(C - c_z)}{(N - n_z)}$ em $L(z; p; q)$,

$$L(z) = \frac{c_z^{c_z} e^{-c_z}}{c_z!} \frac{(C - c_z)^{(C - c_z)} e^{-(C - c_z)}}{(C - c_z)!} = \frac{c_z^{c_z} (C - c_z)^{(C - c_z)} e^{-C}}{c_z! (C - c_z)!}$$

Portanto, encontramos a seguinte razão de verossimilhança para o nosso modelo Poisson descrito em 2.1:

$$LR(z) = \frac{L}{L_0} = \begin{cases} \left(\frac{c_z}{\mu_z}\right)^{c_z} \left(\frac{C - c_z}{C - \mu_z}\right)^{C - c_z} & , \text{ caso } c_z > \mu_z \\ 1 & , \text{ caso o contrário} \end{cases}$$

Para identificar o conglomerado mais verossímil, utilizaremos esta razão maximizada sobre todas as zonas e com a seguinte estatística do teste:

$$T = \max_z LR(z)$$

Normalmente, pelo fato da razão crescer muito rápido, utilizamos seu logaritmo, ou seja, $LLR(z) = \log LR(z)$. Uma vez que a função logaritmo é estritamente crescente, o valor que maximiza $LR(z)$ também maximiza seu logaritmo $LLR(z)$, portanto:

$$LLR(z) = \log \frac{L}{L_0} = \begin{cases} c_z(\log c_z - \log \mu_z) + (C - c_z)[\log(C - c_z) - \log(C - \mu_z)] & , \text{ caso } c_z > \mu_z \\ 0 & , \text{ caso o contrário} \end{cases} \quad (2.3)$$

E, portanto, a estatística do teste:

$$T = \max_z LLR(z) \quad (2.4)$$

Porém esta estatística T possui uma distribuição desconhecida. Este problema é contornado com o uso do Método de Monte Carlo (MMC), processo de simulação que será descrito mais a frente onde será obtida uma distribuição empírica.

2.2 Estatística Scan Temporal

Para a utilização da estatística citada é necessário definir claramente alguns pontos. Primeiramente deve-se ter claro o conceito de “zona”, que é um conjunto conexo de regiões do mapa. Agora, sob o ponto de vista puramente temporal, não estamos interessados em identificar uma área do mapa onde os casos ocorrem com maior probabilidade, mas uma janela temporal dentro da qual o número de casos é maior que o esperado.

Neste trabalho o banco de dados em estudo está dividido em meses ao longo de alguns anos, ou seja, intervalos de tempo. Assim, a zona definida como zona z , antes observada como um conjunto conexo de regiões no espaço, é tida, agora, como um intervalo de tempo contendo todas as regiões do mapa. Ou seja, uma zona que compreende do mês i ao mês j é definida por $z = [t_i, t_j]$.

Assim, seja C_{jt} o número de casos da doença na região j no tempo t , então, o número de casos em um dado mês t é dado por:

$$C^t = \sum_{j=1}^m C_{jt} \quad (2.5)$$

Onde j define as regiões a serem contabilizadas.

Portanto o C_z que efetivamente será utilizado e que contabiliza a incidência total da doença no intervalo de tempo específico $z = [t_i, t_j]$ é definido, de acordo com 2.5, por:

$$C_z = \sum_{t \in z} C^t$$

Onde o número de casos C_z continua definido como uma variável aleatória de distribuição Poisson cujo parâmetro λ_z , que representa o número esperado de casos, é tal que $\lambda = pn_z$, onde n_z é a populações total do mapa (soma das populações de todas as regiões) dentro do intervalo de tempo z . Assim, a estatística do teste se mantém:

$$T = \max_z LLR(z) \quad (2.6)$$

T continua com uma distribuição desconhecida. Problema contornado com a simulação que será descrita mais a frente onde será obtida uma distribuição empírica.

2.3 Estatística Scan Espaço-temporal

Em sua versão espaço-temporal, a estatística scan considera como zona um conjunto conexo de regiões dentro de uma janela temporal delimitada, isto é, o número de casos C_z será contabilizado pelos casos das regiões dentro de z que ocorreram no intervalo de tempo considerado na janela temporal. O mesmo ocorre para efeito de contabilizar n_z e os complementares $C_{\bar{z}}$ e $n_{\bar{z}}$.

2.4 Candidatos a conglomerados

2.4.1 Caso Espacial

No caso espacial, os candidatos a conglomerados são porções do mapa delimitadas geograficamente. Como a janela nesse caso é puramente espacial, em termos temporais a zona z compreende todo o intervalo de tempo e, portanto, o número de casos da zona é a soma dos casos que ocorrem dentro da janela espacial em qualquer instante de tempo. O mesmo vale para a população.

O procedimento constitui em partir de uma região e seguir adicionando, a cada iteração, a região mais próxima. Assim, partindo a região r_1 , fazemos $z_1 = r_1$. Calcula-se então o valor de $T = LLR(z_1)$. Em seguida considera-se $z_2 = (r_1, r_{(2)})$, onde $r_{(2)}$ é a região mais próxima de r_1 . O valor de $LLR(z_2)$ é então comparado ao valor de T , sendo este último substituído por $LLR(z_2)$ caso $LLR(z_2) > T$.

As regiões do mapa vão sendo então adicionadas sequencialmente à zona, na ordem de distância a r_1 , até que se atinja uma zona que contenha, no máximo, 50% da população total ou 50% do total do mapa. Em seguida o procedimento é reiniciado, partindo de cada uma das regiões do mapa. Ao final, o candidato z^* a cluster mais verossímil será aquele cujo valor $LLR(z^*)$ foi o maior observado.

2.4.2 Caso Temporal

No caso temporal, os candidatos a conglomerados são obtidos a partir de um procedimento que varre intervalos de tempo de tamanho variável em busca de conglomerados. Na forma de conjuntos de meses subsequentes¹, buscando nestas zonas evidências da existência de conglomerados de acordo com o valor da estatística LLR obtida.

O processo constitui-se em avaliar, a cada iteração, uma zona $z = [t_i, t_j]$, ou seja, o conjunto de meses subsequentes do i -ésimo ao j -ésimo mês. Como a janela nesse caso é puramente temporal, em termos espaciais a zona z compreende todo o mapa e, portanto, o número de casos da zona é a soma dos casos que ocorrem dentro da janela temporal em qualquer região do mapa. Obviamente, o mesmo vale para a população.

Iniciando o procedimento pela primeira porção de tempo, a primeira zona selecionada será formada somente pelo mês t_1 , ou seja, $z_1 = t_1$. Calcula-se então o valor inicial da estatística T tida por $T = LLR(z_1)$ como visto anteriormente.

Segue-se com o cálculo na próxima região z_2 composta por $z_2 = [t_1, t_2]$ onde os meses t_1 e t_2 são subsequentes. Com os dados da zona z_2 é calculado o $LLR(z_2)$. Como visto, na fórmula 2.6, o valor da estatística T é o valor máximo de $LLR(z)$, assim, se o valor desta zona supera o valor da anterior, esta será considerada, ainda que provisoriamente, o valor da estatística T .

¹Meses conexos ou subsequentes seriam *janeiro/fevereiro*, *abril/maio* ou *julho/agosto/setembro*, por exemplo.

O procedimento segue então adicionando meses subsequentes à janela temporal, até que o intervalo de tempo avaliado $z_n = [t_1, t_n]$ alcance 50% do período de tempo total considerado no estudo. O procedimento é então repetido, partindo inicialmente de cada um dos instantes de tempo. A estatística $LLR(z)$ obtida na zona atual é sempre comparada com o máximo obtido até o momento. O conglomerado mais verossímil será a zona z^* que corresponder ao valor máximo e, assim, teremos nossa estatística de teste T .

2.4.3 Caso Espaço-Temporal

No caso espaço temporal, os candidatos a cluster são avaliados através de um procedimento híbrido. Assim, para cada zona espacial descrita em 2.4.1, consideramos todos os possíveis subintervalos temporais descritos em 2.4.2.

2.5 Simulação de Monte Carlo

Obtido o valor observado da estatística de teste, $T = LLR(z^*)$, o próximo passo consiste na verificação da significância da mesma. Porém para testarmos este conglomerado mais verossímil não conhecemos a distribuição da estatística T sob H_0 . Portanto, faz-se necessário o uso de simulações de Monte Carlo (Metropolis, 1949) para a estimação da distribuição empírica.

A simulação de Monte Carlo consiste essencialmente na replicação dos casos em estudo com base na hipótese nula H_0 descrita na fórmula 2.2. Assim, o número total C de casos é fixado enquanto o número de casos em cada instante de tempo é distribuído de modo aleatório sob H_0 .

Portanto, a distribuição dos casos C_z mantém-se no formato Poisson, porém com a probabilidade esperada para os casos $p = \frac{C}{N}$. O número esperado de casos em uma região em um dado instante de tempo t é proporcional à sua população naquele instante de tempo. Ou seja, $\mu_z = pn_z$.

$$C_z \sim \text{Poisson}(\lambda_z = \frac{C}{N}n_z)$$

Assim, geram-se C casos em cada réplica do mapa, distribuídos aleatoriamente ao longo do tempo e do espaço. Essa distribuição é replicada milhares de vezes, e para cada replicação executa-se o procedimento scan para se obter o cluster mais verossímil z^* . Assim, para cada replicação, obtemos o valor da estatística T sob H_0 , uma vez que os dados foram gerados sob a hipótese nula. O conjunto destas estatísticas empíricas nos gera uma distribuição empírica da estatística de teste.

Partindo desta distribuição empírica obtida podemos calcular o p-valor. O p-valor é definido como a probabilidade de obtermos uma estatística de teste igual ou mais extrema que a observada em uma dada amostra quando assumimos verdadeira a hipótese nula, ou seja, é favorável um p-valor significativamente baixo no caso em estudo para que seja rejeitada com certa segurança a hipótese nula acerca dos dados em estudo.

Com nível de significância 0,05 a hipótese nula, de que a zona não é um conglomerado, apenas é rejeitada caso:

$$P(E > T) \leq 0,05$$

onde E possui distribuição empírica gerada via Monte Carlo. Pode-se inferir, nesta situação, que o conglomerado encontrado é significativo.

Capítulo 3

O Banco de dados

Este capítulo discutirá a respeito das bases de dados utilizadas neste trabalho. Serão discutidos tanto problema em questão quanto a base de dados aplicada e estudada com o software *SatScan* no próximo capítulo.

3.1 Problema em questão

A técnica Scan descrita minuciosamente no capítulo anterior respalda o procedimento que será implementado e aplicado na base de dados mais a frente com auxílio do software SatScan. Temos o objetivo de detectar um conglomerado temporal de casos de dengue no período de 2007 a 2011 focando no julgamento da hipótese proposta de sazonalidade da doença.

Esta técnica Scan, assim como o software *SatScan*, já foi aplicada à epidemiologia anteriormente (Nagarwalla, 1995) no estudo da leucemia em Nova York no qual os resultados foram obtidos rapidamente e satisfatoriamente onde concluiu-se que a técnica é ótima para a detecção de clusters de doenças em formação ou ascensão. Houve também um outro estudo mais recente (Kulldorff, 2001) que focou-se no crescimento do câncer no Novo México e chegou a conclusões similares que podem ajudar na adoção de medidas preventivas e mais direcionadas a respeito da doença.

3.2 Os dados e o DataSUS

A base de dados que será utilizada neste trabalho foi extraída do domínio do DataSUS (2012). DataSus é o domínio do departamento de Sistema Único de Saúde do nosso país que disponibiliza informações que podem servir para subsidiar análises objetivas da situação sanitária, tomadas de decisão baseadas em evidências e elaboração de programas de ações de saúde.

Ano	Número de casos
2007	500.339
2008	559.895
2009	386.133
2010	984.662
2011	732.160
2012	Indisponível

Tabela 3.1: Número de casos de dengue por ano no Brasil conforme o site do DataSus

A informação é fundamental para a democratização da Saúde e o aprimoramento de sua gestão. A informatização das atividades do Sistema Único de Saúde (SUS), dentro de diretrizes tecnológicas adequadas, é essencial para a descentralização das atividades de saúde e viabilização do Controle Social sobre a utilização dos recursos disponíveis. Para alcançar tais objetivos foi criado o Departamento de Informática do SUS – DataSUS. A partir de 2011 o DataSUS passa a integrar a Secretaria de Gestão Estratégica e Participativa, conforme Decreto Nº 7.530 de 21 de julho de 2011 que trata da Estrutura Regimental do Ministério da Saúde.

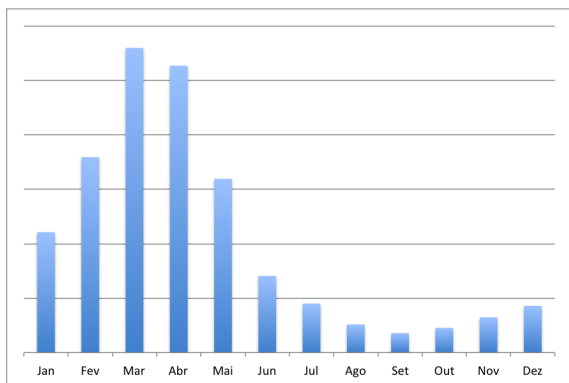
O banco de dados consiste em uma série temporal de casos da doença. De modo sucinto, uma série temporal é qualquer conjunto de observações ordenadas no tempo, ou seja, consiste em um conjunto de observações geradas sequencialmente no tempo. Estes números de incidência da doença estão organizados por município classificados por mês e ano a partir do ano de 2007, e contém apenas os casos confirmados.

De forma mais geral, podemos observar pela tabela 3.1 tanto a instabilidade da doença ao longo dos anos quanto o seu crescimento firme. A base de dados organiza-se como na tabela, porém mais minuciosa com os casos divididos por cada município do Brasil ao longo dos meses.

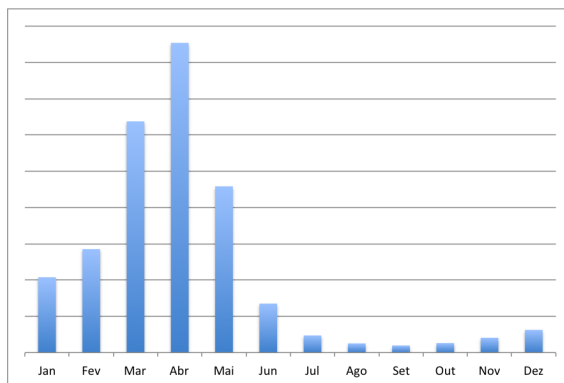
Para a identificação de uma epidemia é necessário conhecer a frequência precedente da doença. Com cunho ilustrativo, são apresentados na figura 3.1 histogramas, obtidos através do software R (2012), para os períodos da base de dados posta em estudo. É uma maneira simples e útil de se ter uma idéia da curva epidêmica, consistindo na representação gráfica das frequências mensais da doença num eixo de coordenadas, no qual o eixo horizontal representa o tempo e o vertical, as frequências.

Observamos que as médias são realmente maiores que a mediana indicando uma distribuição altamente assimétrica. Repare que a curva descendente representa a fase de esgotamento da epidemia conforme o ano passa.

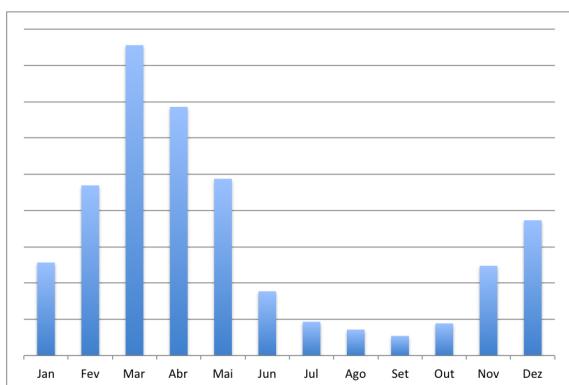
Assim, podemos inferir que há estados com grande número de casos e outros com pequeno número de casos, o que seria normal, a princípio, pois alguns estados têm população muito pequena (esperando-se baixa incidência) enquanto outros têm população grande (esperando-se alta incidência). Mas este fato pode indicar também a existência de clusters espaciais ou espaço-temporais. Essa hipótese será verificada no andar do trabalho.



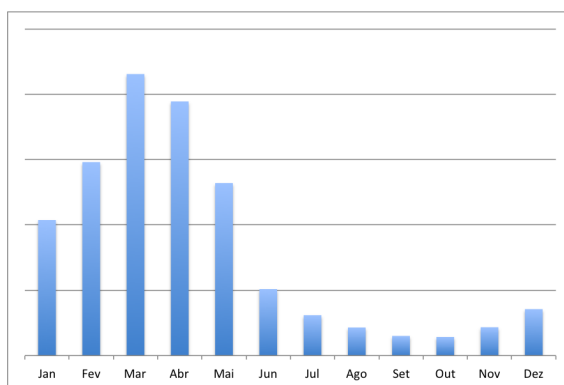
(a) 2007



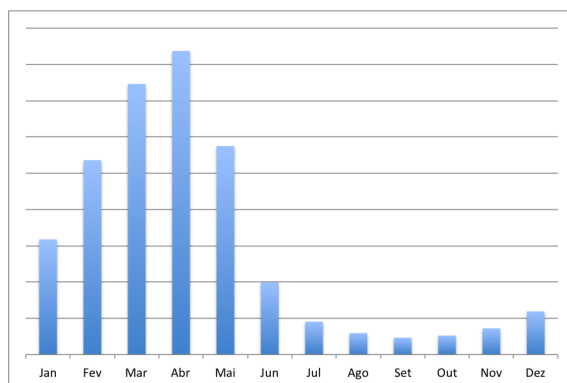
(b) 2008



(c) 2009



(d) 2010



(e) 2011

Figura 3.1: Histogramas, com fim ilustrativo e exploratório, para cada período do banco de dados. A curva epidêmica tem usualmente distribuição assimétrica, repare na absurda proporção de casos na porção inicial de cada ano.

Capítulo 4

Discussão dos Resultados

Neste capítulo serão discutidos os resultados das análises realizadas. O capítulo está dividido em três seções, a primeira consiste na apresentação da saída bruta obtida com o software. A segunda parte traz uma discussão mais profunda dos resultados e estatísticas obtidas. A terceira e última parte desta discussão faz considerações finais e coloca a sua importância no papel do combate à dengue.

4.1 Análise Puramente Temporal

Nesta seção será apresentada a análise temporal realizada assim como os resultados que foram obtidos.

4.1.1 Apresentação do output do software

Nesta seção serão apresentados os resultados fornecidos pelo software para os anos de 2007 a 2011, que serão analisados na seção subsequente. Ao final é apresentada uma análise do intervalo completo de 2007 a 2011. O output completo utilizado para análise pode ser encontrado no *Apêndice A* no final deste trabalho.

O seguinte output refere-se ao ano de 2007, de primeiro de janeiro a 31 de dezembro.

SUMMARY OF DATA

Study period.....: 2007/1/1 to 2007/12/31

MOST LIKELY CLUSTER

1.Location IDs included.: All
Time frame.....: 2007/1/1 to 2007/5/31
Number of cases.....: 397442
Expected cases.....: 206977.09
Annual cases / 100000.: 507.8
Observed / expected...: 1.92
Relative risk.....: 5.47

Log likelihood ratio...: 151504.434202
P-value.....: 0.001

Nesta primeira parte do output são fornecidas informações importantes como o cluster mais verossímil, que é a janela temporal buscada neste trabalho, assim como a estatística do teste $LLR(z)$ e informações importantes a respeito deste cluster como o risco relativo e uma razão entre os casos observados e esperados, que são medidas importantes e muito utilizadas na epidemiologia. O P-valor, probabilidade de obtermos uma estatística de teste igual ou mais extrema que a observada quando assumimos H_0 , também é fornecido. Observe, ainda, que o software inclui no cluster todas as localizações espaciais, indicando que a análise foi puramente temporal.

PARAMETER SETTINGS

Analysis

Type of Analysis : Retrospective Purely Temporal
Probability Model : Discrete Poisson
Scan for Areas with : High Rates
Time Aggregation Units : Month
Time Aggregation Length : 1

Output

Results File : /Users/rafaellewkowicz/Downloads/dengue/resultados/Resultados 2007

Observando a segunda parte do output, notamos que o banco de dados foi varrido pelo algoritmo de mês em mês em busca de zonas com alta concentração de casos.

Repare que o algoritmo foi executado com uma opção que nos permite ignorar casos e controles fora do período de análise estipulado, assim, podemos trabalhar com apenas uma base de dados condensada com todos os casos de dengue de primeiro de janeiro de 2007 até 31 de dezembro de 2011, ou em diante, caso necessário futuramente.

Data Checking

Temporal Data Check : Ignore cases and controls that are outside the specified temporal study period.

Temporal Window

Maximum Temporal Cluster Size : 50 percent of study period

Inference

```

-----
P-Value Reporting           : Default Combination
Adjusting for More Likely Clusters : No
Number of Replications      : 999

```

Run Options

```

-----
Processor Usage   : All Available Processors
Logging Analysis  : Yes
Suppress Warnings : No
-----

```

```

Total Running Time : 10 seconds
Processor Usage     : 2 processors

```

Por último existem informações a respeito da simulação de Monte Carlo e do processamento do software durante a análise.

4.1.2 Análise e Discussão

O algoritmo foi executado através do SatScan versão 9.1.1. O processamento ocorreu em uma máquina com sistema operacional OS X 10.8.2 com dois núcleos de 2GHz e 4GB de memória RAM. Para cada ano o algoritmo precisou de cerca de 13 segundos de processamento para ser completado.

Sabemos que a dengue é uma doença que está claramente fora de controle necessitando de constante atenção dos agentes de saúde em relação ao seu combate. A tabela 4.1 nos dá uma visão inicial mais clara em relação à magnitude do problema. Repare que as razões e os riscos relativos são referentes aos clusters temporais encontrados, e não ao ano inteiro.

Ano	Razão	Risco Relativo
2007	1.92	5.47
2008	2.42	8.02
2009	1.97	3.74
2010	1.95	5.92
2011	2.07	8.25
Geral	3.06	3.75

Tabela 4.1: Razão entre os valores observados e esperados e o risco relativo para a doença para cada cluster.

Claramente podemos notar que a incidência, nos clusters encontrados ao longo dos anos, é aproximadamente três vezes maior que a esperada, ou seja, a população encontra-se três vezes mais exposta à doença do que deveria estar. Observando cada ano separadamente vê-se que há, em média, duas vezes mais casos que o esperado e em 2008 temos praticamente duas vezes e meia mais casos do que deveriam existir o que mostra que, nas janelas temporais referentes

Ano	Cluster	$LLR(z)$	P-valor
2007	Janeiro a Maio	$1,52 \cdot 10^5$	0.001
2008	Fevereiro a Maio	$2,59 \cdot 10^5$	0.001
2009	Fevereiro a Maio	$8,16 \cdot 10^4$	0.001
2010	Janeiro a Maio	$3,20 \cdot 10^5$	0.001
2011	Fevereiro a Junho	$2,96 \cdot 10^5$	0.001
2007/2011	Janeiro a Maio de 2010	$4,06 \cdot 10^5$	0.001

Tabela 4.2: As janelas mais verossímeis, as estatísticas do teste e os P-valores relativos para o caso temporal.

aos clusters, a doença está claramente fora de controle e, com tamanho impacto em território nacional, podemos caracterizá-la sem relutância como apresentando um perfil endêmico¹.

Em seguida, podemos observar uma medida interessante apresentada; o risco relativo. O risco relativo é claramente maior que 1, isto é, o risco da doença é maior em indivíduos expostos do que em indivíduos não expostos (indicando uma relação da exposição do indivíduo à doença em pauta). Notamos um risco de 8.02 em 2008, ou seja, existia, para a população exposta, um risco 8 vezes maior de adquirir a doença neste ano, o que faz pleno sentido uma vez que 2008 foi o ano em que ocorreu uma gigantesca epidemia no Brasil superando, em número de vítimas fatais, a epidemia de 2002 onde 91 pessoas morreram. Somente no estado do Rio de Janeiro foram registrados 250 mil casos e por volta de 200 mortes em todo o estado, sendo 100 mortes e 125 mil casos somente na capital. Devido ao grande impacto da epidemia de 2008 o governo voltou com um combate intensivo diminuindo o risco consideravelmente para 3.74 em 2009.

Seguindo, a segunda parte da análise (tabela 4.2) nos traz informações como a estatística de Log-Verossimilhança e os conglomerados de maior incidência com os P-valores relativos às simulações de Monte Carlo.

Utilizada a simulação de Monte Carlo para estimação da distribuição empírica, foram feitas 999 replicações sob a hipótese de que possíveis conglomerados de casos ocorrem por mero acaso, ou seja, a hipótese H_0 de que a probabilidade de que um indivíduo no conglomerado z venha a ser um caso é igual à probabilidade de que um indivíduo fora de z venha a ser um caso.

Notamos um fato interessante ao testar as estatísticas $LLR(z)$ com sua distribuição empírica a um nível de significância 0.05: a probabilidade de obtermos uma estatística de teste igual ou mais extrema que a observada é muito baixa quando assumimos verdadeira a hipótese H_0 , ou seja, o p-valor significativamente baixo no caso comprova com grande segurança a hipótese acerca dos dados em estudo validando que os conglomerados encontrados não ocorrem por mero acaso.

Observando os conglomerados obtidos, note que existe um padrão nos resultados, indicando que há uma concentração de casos no primeiro semestre de cada ano, mais precisamente de janeiro a maio. O conglomerado mais intenso encontrado, ao se estudar o intervalo de tempo completo, está localizado de janeiro a maio de 2010 o que pode ser explicado pela epidemia

¹Quando uma grande incidência está acima da esperada e a doença se dissemina rapidamente num segmento demográfico da população humana. (Guimarães Jr., 2001)

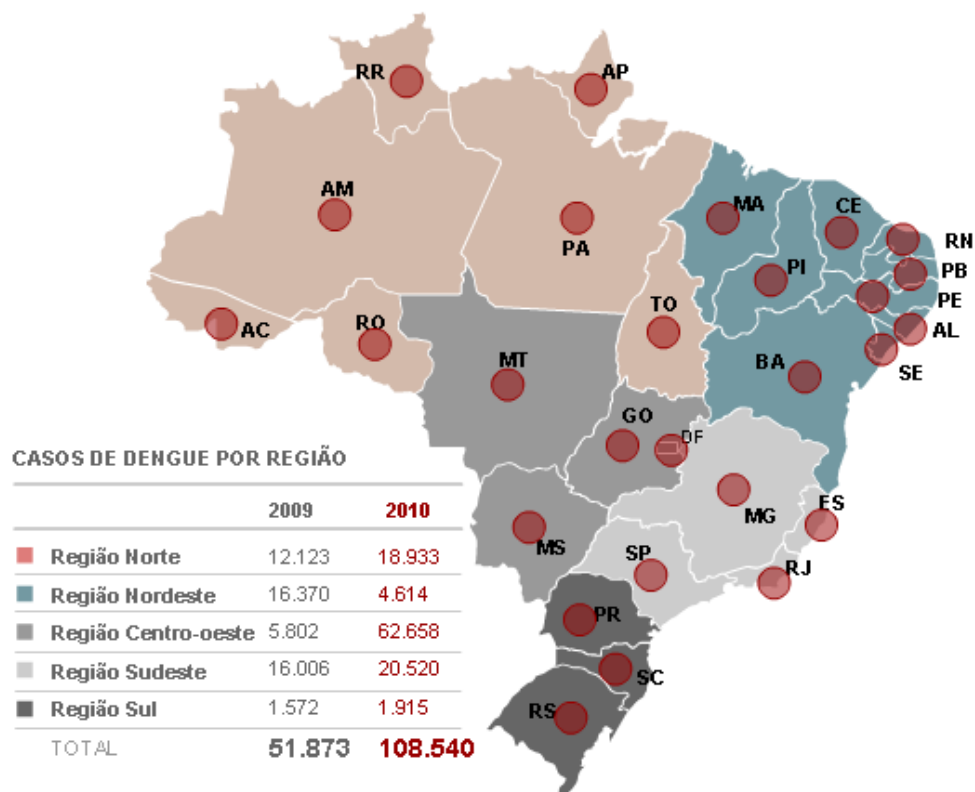


Figura 4.1: O crescimento de 2009 para 2010. Número de casos encontrados até fevereiro destes anos. Fonte: Ministério da Saúde

ocorrida no início deste ano onde ocorreram mais de 108 mil registros de casos somente até a metade do mês de janeiro, o que significa, segundo o ministério da saúde, um repentino aumento de 109% em relação ao ano anterior dando desfecho ao crescimento que vinha ocorrendo desde 1995 como ilustrado na figura 1.1 no início deste trabalho. O crescimento ocorrido dentro desta janela é comparado ano anterior na figura 4.1. Um fato importante que observamos nesta análise é o de que os surtos estão explicitamente ligados ao período de chuvas.

Outro fato interessante é o de que, passando por este conglomerado até 16 de outubro de 2010, 592 pessoas morreram no país por causa da doença, contra 298 em todo o ano de 2009. Em setembro, o número de mortes causadas pela doença já era recorde neste ano, superando a marca de 2008, quando a doença fez 478 vítimas. Dos 936.260 casos de dengue clássica já classificados no país até esse período, 14.342 foram classificados como graves o que pode ser justificado pelo fato de que o país está vivendo, desde os anos 90, surtos repetidos de dengue, que eram causados pelos três tipos do vírus que circulam no Brasil. Isso aumenta as chances de uma pessoa se infectar novamente e, até mesmo, desenvolver as formas graves da doença.

De acordo com o ministério da saúde:

Outro fator que tem contribuído para o aumento no número de casos de dengue neste ano é o retorno do sorotipo viral conhecido como “DEN-1”, que circulou com maior intensidade na década de 90. Esse sorotipo viral que voltou a predominar em alguns estados, e é um fator que pode contribuir para que se estabeleça uma epidemia da doença no Brasil.

Ministério da Saúde, Secretaria de Vigilância Em Saúde.

Segundo Giovanini Coelho, coordenador geral do Programa Nacional de Controle da Dengue do Ministério da Saúde, todas as epidemias que já existiram se deveram ao aparecimento de um novo sorotipo, ou ao retorno de um sorotipo viral que não circulava há muito tempo.

Todas estas concentrações apresentadas na tabela 4.2 ocorrem no primeiro semestre de cada ano e são significativas, o que significa que elas não ocorrem por mero acaso e estão possivelmente ligadas à sazonalidade.

4.2 Análise Espaço-Temporal

Para a análise espacial, algoritmo foi executado através do SatScan versão 9.1.1. em um computador AMD 6100 com poder de processamento de 6 núcleos reais de 3GHZ e 8GB de memória RAM. O sistema operacional utilizado desta vez foi o Windows 7 64 bits. Para cada ano o algoritmo precisou de 30 a 40 minutos de processamento para ser completado, sendo que na análise para o período completo foram precisas mais de 17 horas de processamento.

As coordenadas dos clusters espaço-temporais geradas pelo SatScan foram processadas no MatLab tornando possível que fossem gerados mapas dando uma visualização mais consistente dos clusters. A figura 4.2 representa o cluster obtido para o período total de estudo, de 2007 a 2011, sendo que todos os clusters apresentados neste trabalho foram tidos como significativos e quase a totalidade com um p-valor menor que $0.001 \cdot 10^{-15}$.



Figura 4.2: O Cluster Espaço-Temporal mais significativo no período total de estudo.

Observe que a tabela 4.3 nos traz, para cada cluster espaço-temporal, informações como a estatística de Log-Verossimilhança, os conglomerados de maior incidência e os P-valores relativos às simulações de Monte Carlo. Segue, para uma melhor visão geral dos resultados, uma lista dos clusters obtidos. Note que alguns são formados por todas as regiões, assim, não faz sentido colocá-los neste trabalho²;

- Em 2007, foram obtidos 4 clusters significativos sendo o 2º formado por todas as regiões.
- Em 2008, foram obtidos 2 clusters significativos sendo o 2º formado por todas as regiões.
- Em 2009, foram obtidos 7 clusters significativos sendo o 3º formado por todas as regiões.
- Em 2010, foram obtidos 2 clusters significativos sendo o 2º formado por todas as regiões.
- Em 2011, foram obtidos 3 clusters significativos sendo o 1º formado por todas as regiões.



Figura 4.3: Clusters espaço-temporais mais significativos de 2007.

²Note que os clusters puramente temporais, obviamente, também foram omitidos uma vez que são constituídos pelo mapa inteiro.

Ano	Cluster Espaço-Temporal	$LLR(z)$	P-valor
2007	Janeiro a Maio	$2,45 \cdot 10^5$	$<0.001 \cdot 10^{-15}$
	Janeiro a Maio*	$1,51 \cdot 10^5$	$<0.001 \cdot 10^{-15}$
	Março a Maio	$3,54 \cdot 10^4$	$<0.001 \cdot 10^{-15}$
	Março a Junho	$2,15 \cdot 10^5$	$<0.001 \cdot 10^{-15}$
2008	Fevereiro a Maio	$4,24 \cdot 10^5$	$<0.001 \cdot 10^{-15}$
	Fevereiro a Maio*	$2,59 \cdot 10^5$	$<0.001 \cdot 10^{-15}$
2009	Fevereiro a Maio	$1,87 \cdot 10^5$	$<0.001 \cdot 10^{-15}$
	Janeiro a Dezembro	$1,56 \cdot 10^5$	$<0.001 \cdot 10^{-15}$
	Fevereiro a Maio*	$8,15 \cdot 10^4$	$<0.001 \cdot 10^{-15}$
	Fevereiro a Março	82,67	$<0.001 \cdot 10^{-15}$
	Abril	68,95	$<0.001 \cdot 10^{-15}$
	Janeiro a Maio	53,29	$0.0056 \cdot 10^{-15}$
	Agosto a Setembro	19,90	0.0012
2010	Janeiro a Maio	$5,51 \cdot 10^5$	$<0.001 \cdot 10^{-15}$
	Janeiro a Maio*	$3,19 \cdot 10^5$	$<0.001 \cdot 10^{-15}$
2011	Fevereiro a Junho*	$2,96 \cdot 10^5$	$<0.001 \cdot 10^{-15}$
	Março a Junho	$2,12 \cdot 10^5$	$<0.001 \cdot 10^{-15}$
	Fevereiro a Abril	$1,29 \cdot 10^5$	$<0.001 \cdot 10^{-15}$
2007/2011	Janeiro a Maio de 2010	$6,63 \cdot 10^5$	$<0.001 \cdot 10^{-15}$
	Janeiro a Maio de 2010*	$4,06 \cdot 10^5$	$<0.001 \cdot 10^{-15}$

Tabela 4.3: As janelas espaço-temporais mais verossímeis, as estatísticas do teste e os P-valores relativos. * Repare que os clusters marcados englobam o mapa completo.

Quanto ao principal cluster de 2007, figura 4.5(a), observa-se que os principais estados afetados são o Paraná, São Paulo e, principalmente, o Mato Grosso do Sul. Em 2007, segundo o ministério da saúde, o Mato Grosso do Sul representou 67% das notificações da região centro-oeste (74.902 casos) e apresentou a maior incidência (3.213 casos por 100.000 habitantes).

Essa região apresentou o maior aumento no número de casos comparado-se com o mesmo período de 2006 (807%), em virtude das transmissões ocorridas em vários municípios do Estado do Paraná. A maior transmissão de dengue ocorreu nos municípios com menos de 100.000 habitantes.

Comparando o ano de 2007 com o ano de 2006, verifica-se uma redução de 46,7% no número de casos notificados em Goiás e um aumento nos demais estados sendo de 373% Mato Grosso do Sul, de 35% Mato Grosso e 115% no Distrito Federal.

Já o terceiro cluster de 2007, figura 4.5(b), engloba os estados Bahia, Minas Gerais, Espírito Santo e Rio de Janeiro. O Estado do Rio de Janeiro notificou 63.180 casos, dos quais 39% foram notificados na capital (24.594 casos), 12% (7.636 casos) em Niterói e 5% (3.187 casos) em São Gonçalo. O Estado de Minas Gerais notificou 41.785 casos de dengue dispersos por diversos municípios. O Estado do Espírito Santo notificou 11.001 casos de dengue (313 casos por 100.000 habitantes), o que representa uma redução de 23% quando comparado ao ano de 2006. Informações conforme o Informe Epidemiológico da Dengue, Janeiro a Dezembro de 2007, do ministério da saúde.

O cluster mais significativo de 2008 nos mostra uma janela circular que envolve todo o litoral leste do país vindo do Rio Grande passando pelos estados do Piauí e Bahia, e finaliza envolvendo o estado do Rio de Janeiro. Observe a figura 4.4.



Figura 4.4: O Cluster Espaço-Temporal mais significativo de 2008.

O cluster de 2009 engloba fortemente Bahia, Minas Gerais e Espírito Santo, o que faz sentido uma vez que quase 40% dos casos de dengue hemorrágica ocorreram nestes estados em 2009. Houve uma queda do número de casos em todos os estados, como podemos observar na tabela 3.1, exceto na Bahia e no Piauí, sendo que no estado da Bahia houve um aumento de 227% em 2009. O segundo cluster mais significativo apontou maior atividade nos estados de Mato Grosso, Mato Grosso do Sul e Goiás. A região com maior número de óbitos em 2009 foi o Nordeste, com 115 registros só na Bahia foram 66 mortes.

Ainda em 2009, o estado do Paraná mostrou-se fortemente afetado pela doença com incidência de 7,33/100.000 habitantes, que torna-se alta ao ponderarmos pelas suas pequenas dimensões. Os municípios presentes no cluster representado em 4.5(c) são os seguintes: Bela Vista do paraíso, Sertanópolis, Primeiro de maio, Rancho Alegre D'Oeste, Quarto Centenário, Goioerê, Juranda, Boa esperança, Janiópolis, Ubitatã, Moreira Sales, Campina da Lagoa, Formosa do Oeste, Nova Aurora, Mariluz, Farol, Anahy, Mamborê, Iracema do Oeste, Jesuítas, Iguatu, Tuneiras do Oeste, Cafelândia, Nova, Cantu, Assis Chateaubriand, Guaraci e Jataizinho. Dentre destes, o município de Toledo apresenta a maior incidência de dengue hemorrágica do estado, sendo de 77,58 para cada 100.000 habitantes, e também apresenta a maior concentração de número de casos confirmados (251). Já Assis Chateaubriand, com 875,87 casos para cada 100.000 habitantes, apresenta a segunda maior incidência do estado neste ano, sendo Icaraíma a primeira com 1.058,01 casos para cada 100.000 habitantes. Apesar da forte presença da doença, não houve óbitos no Paraná neste ano.

Em 2010 e 2011, fica claro no mapa que a doença continua bastante alastrada pelo país. O clusters principais e secundários de 2010 e 2011 abrangem quase todo o território nacional, sendo que 2011 é o único período em que ocorre o fato de um cluster primário mais significativo capturar o mapa inteiro.

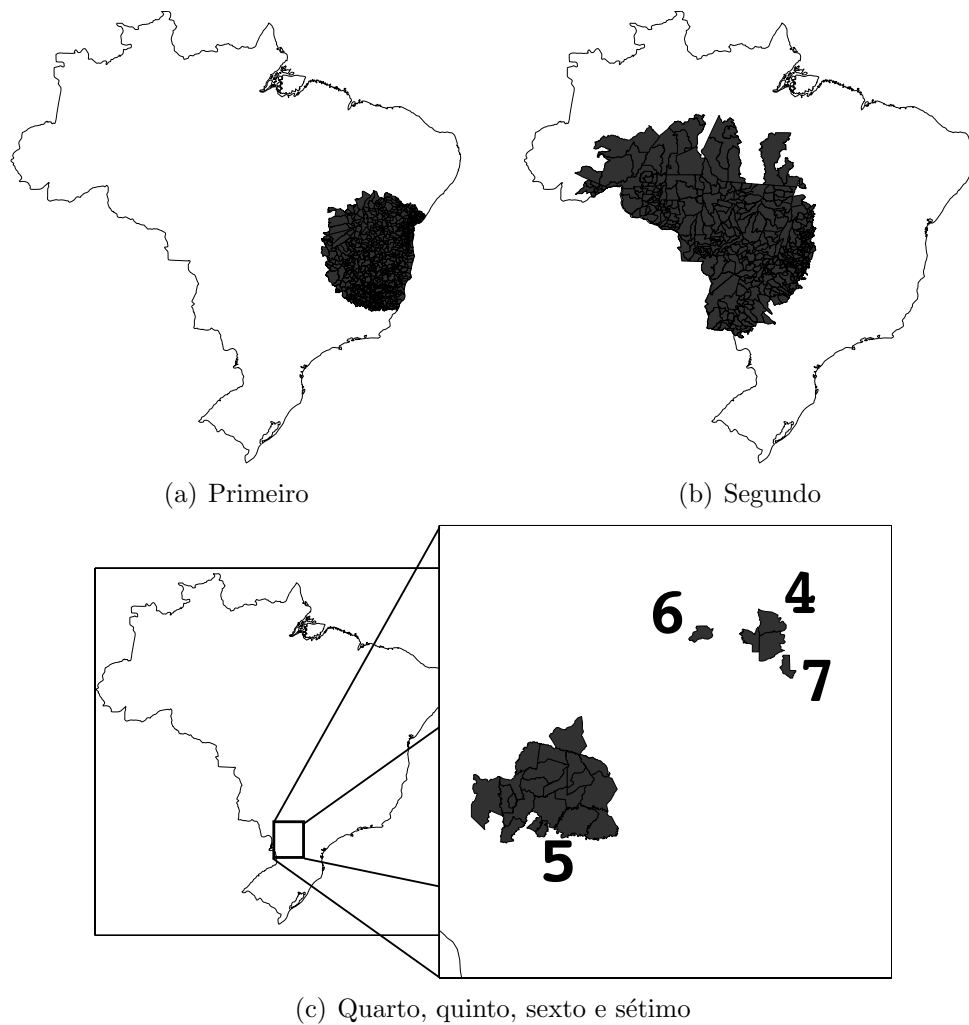


Figura 4.5: Os clusters espaço-temporais mais significativos de 2009.

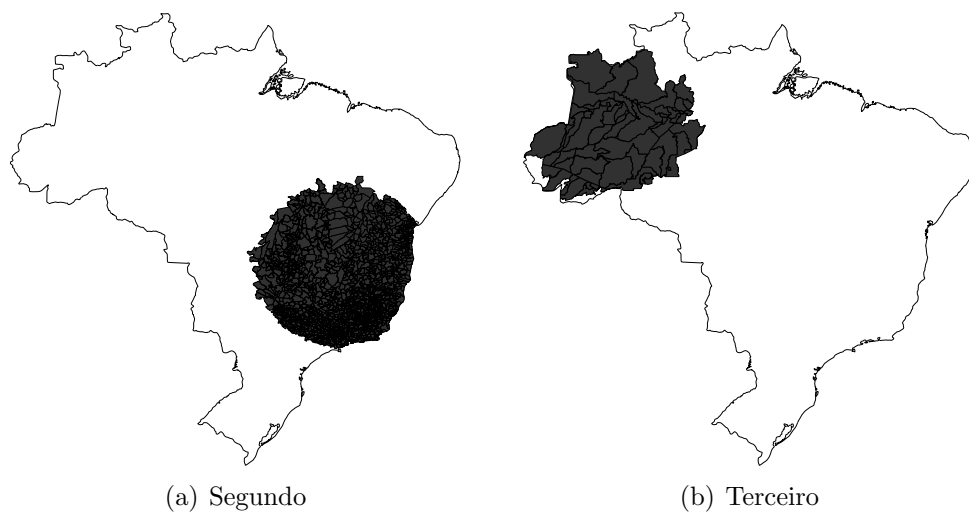


Figura 4.6: Os clusters espaço-temporais mais significativos de 2011, sendo que o primeiro é constituído de todas as regiões.

Observando o mapa do intervalo completo de 2007 a 2011, veja figura 4.2, podemos concluir algo próximo do que foi observado estudando os anos separadamente: alguns focos

seriam as regiões Centro-Oeste, Sudeste (observe a figura 4.6(a)), parte da região norte, o norte do Paraná e o estado da Bahia que são regiões que estão sob o constante ataque do vírus, porém o perfil endêmico da doença está verdadeiramente representado uma vez que todos os anos possuem, em algum momento, um cluster que aponta infestação no mapa inteiro. As altas temperaturas, grande volume de chuvas, o desmatamento e a constante urbanização e consequente aumento da densidade populacional podem explicar a epidemia nestas regiões.



Figura 4.7: O Cluster Espaço-Temporal mais significativo de 2010.

Capítulo 5

Considerações finais e trabalhos futuros

5.1 Considerações finais

Neste trabalho, verificamos e validamos a hipótese de sazonalidade na ocorrência de casos de dengue no Brasil utilizando a estatística de varredura proposta por Kulldorff para a identificação de clusters temporais. Uma vez que a distribuição dessa estatística sob H_0 é desconhecida, foram feitas diversas simulações com base nesta hipótese construindo-se uma distribuição de $LLR(z)$ para testar a significância dos conglomerados encontrados. Concluiu-se que a hipótese H_0 foi rejeitada nos mostrando que os casos de dengue não ocorrem ao mero acaso. Temos, assim, como enxergar melhor o funcionamento desta doença que vem ocorrendo de forma tão impactante no nosso país há mais de duas décadas.

Diante deste quadro que estamos observando, percebemos que as autoridades sanitárias devem lançar uma grande campanha de conscientização e combate à dengue, principalmente neste período do primeiro semestre de cada ano, de janeiro a julho, veiculada por rádio, jornais, televisão e revistas de todo o território nacional. Com base nas informações e métodos apresentados neste trabalho é possível também um estudo prospectivo para que seja criado um projeto preventivo de combate a dengue, principalmente em períodos de surto iminente.

Contudo, a diversidade dos tipos de criadouros do mosquito, e o seu número, tornam o combate a dengue um grande desafio de saúde pública. Portanto, o controle desta doença depende não apenas de uma ação intensa do governo e federal, como também do esforço de todos, incluindo comunidades em geral, municípios.

É preciso não só uma ação do governo como também que toda população brasileira se conscientize que estamos enfrentando epidemias contínuas e os casos de dengue não param de subir, inclusive os de dengue hemorrágica, que são os mais graves e muitas vezes levam a óbito.

5.2 Trabalhos futuros

Uma vez que temos disponível o banco de dados do DataSus, que é bem amplo e atualizado com certa regularidade, torna-se possível, uma vez que os métodos já estão implementados, fazer uma atualização do banco de dados para que se possa seguir com o acompanhamento da doença.

Uma análise prospectiva também é possível e possibilita, inclusive, além do monitoramento da doença, uma previsão acerca do seu desenvolvimento permitindo um melhor planejamento para enfrentar quadros mais graves que podem estar próximos.

Outros métodos de detecção de clusters poderiam também ser utilizados, incluindo métodos de geometria arbitrária e métodos não circulares, no caso espaço-temporal.

Apêndice A

Output do Software

Neste capítulo estão os resultados brutos do software utilizado SatScan. O algoritmo foi rodado para os anos de 2007, 2008, 2009, 2010 e 2011 além de uma análise completa de todos os anos para uma análise exploratória.

A.1 Ano de 2007

```
-----  
SaTScan v9.1.1  
-----
```

Program run on: Sat Jan 5 15:25:49 2013

Retrospective Purely Temporal analysis
scanning for clusters with high rates
using the Discrete Poisson model.

SUMMARY OF DATA

Study period.....: 2007/1/1 to 2007/12/31
Total population.....: 189343660
Total number of cases.....: 500339
Annual cases / 100000.....: 264.4

MOST LIKELY CLUSTER

1.Location IDs included.: All
Time frame.....: 2007/1/1 to 2007/5/31
Number of cases.....: 397442
Expected cases.....: 206977.09
Annual cases / 100000.: 507.8
Observed / expected...: 1.92

Relative risk.....: 5.47
Log likelihood ratio..: 151504.434202
P-value.....: 0.001

PARAMETER SETTINGS

Input

Case File : /Users/rafaellewkowicz/Downloads/dengue/dengue.cas
Population File : /Users/rafaellewkowicz/Downloads/dengue/dengue.pop
Time Precision : Month
Start Time : 2007/1/1
End Time : 2007/12/31

Analysis

Type of Analysis : Retrospective Purely Temporal
Probability Model : Discrete Poisson
Scan for Areas with : High Rates
Time Aggregation Units : Month
Time Aggregation Length : 1

Output

Results File : /Users/rafaellewkowicz/Downloads/dengue/resultados/Resultados 2007

Data Checking

Temporal Data Check : Ignore cases and controls that are outside the specified temporal study period.

Temporal Window

Maximum Temporal Cluster Size : 50 percent of study period

Space And Time Adjustments

Temporal Adjustment : None
Adjust for known relative risks : No

Inference

P-Value Reporting : Default Combination
Adjusting for More Likely Clusters : No
Number of Replications : 999

Additional Output

Report Critical Values : No
Report Monte Carlo Rank : No

Run Options

Processor Usage : All Available Processors
Logging Analysis : Yes
Suppress Warnings : No

Program completed : Sat Jan 5 15:25:59 2013
Total Running Time : 10 seconds
Processor Usage : 2 processors

A.2 Ano de 2008

SaTScan v9.1.1

Program run on: Sat Jan 5 15:28:17 2013

Retrospective Purely Temporal analysis
scanning for clusters with high rates
using the Discrete Poisson model.

SUMMARY OF DATA

Study period.....: 2008/1/1 to 2008/12/31
Total population.....: 189686018
Total number of cases.....: 559895
Annual cases / 100000.....: 294.6

MOST LIKELY CLUSTER

1.Location IDs included.: All
Time frame.....: 2008/2/1 to 2008/5/31
Number of cases.....: 447026
Expected cases.....: 185027.48
Annual cases / 100000.: 711.7
Observed / expected...: 2.42
Relative risk.....: 8.02
Log likelihood ratio...: 258845.450213

P-value.....: 0.001

PARAMETER SETTINGS

Input

Case File : /Users/rafaellewkowicz/Downloads/dengue/dengue.cas
Population File : /Users/rafaellewkowicz/Downloads/dengue/dengue.pop
Time Precision : Month
Start Time : 2008/1/1
End Time : 2008/12/31

Analysis

Type of Analysis : Retrospective Purely Temporal
Probability Model : Discrete Poisson
Scan for Areas with : High Rates
Time Aggregation Units : Month
Time Aggregation Length : 1

Output

Results File : /Users/rafaellewkowicz/Downloads/dengue/resultados/Resultados 2008.txt

Data Checking

Temporal Data Check : Ignore cases and controls that are outside the specified temporal study period.

Temporal Window

Maximum Temporal Cluster Size : 50 percent of study period

Space And Time Adjustments

Temporal Adjustment : None
Adjust for known relative risks : No

Inference

P-Value Reporting : Default Combination
Adjusting for More Likely Clusters : No
Number of Replications : 999

Additional Output

Report Critical Values : No

Report Monte Carlo Rank : No

Run Options

Processor Usage : All Available Processors
Logging Analysis : Yes
Suppress Warnings : No

Program completed : Sat Jan 5 15:28:29 2013
Total Running Time : 12 seconds
Processor Usage : 2 processors

A.3 Ano de 2009

SaTScan v9.1.1

Program run on: Sat Jan 5 15:29:41 2013

Retrospective Purely Temporal analysis
scanning for clusters with high rates
using the Discrete Poisson model.

SUMMARY OF DATA

Study period.....: 2009/1/1 to 2009/12/31
Total population.....: 191440641
Total number of cases.....: 386133
Annual cases / 100000.....: 201.8

MOST LIKELY CLUSTER

1.Location IDs included.: All
Time frame.....: 2009/2/1 to 2009/5/31
Number of cases.....: 249867
Expected cases.....: 126967.60
Annual cases / 100000.: 397.2
Observed / expected...: 1.97
Relative risk.....: 3.74
Log likelihood ratio...: 81559.533036
P-value.....: 0.001

PARAMETER SETTINGS

Input

Case File : /Users/rafaellewkowicz/Downloads/dengue/dengue.cas
Population File : /Users/rafaellewkowicz/Downloads/dengue/dengue.pop
Time Precision : Month
Start Time : 2009/1/1
End Time : 2009/12/31

Analysis

Type of Analysis : Retrospective Purely Temporal
Probability Model : Discrete Poisson
Scan for Areas with : High Rates
Time Aggregation Units : Month
Time Aggregation Length : 1

Output

Results File : /Users/rafaellewkowicz/Downloads/dengue/resultados/Resultados 2009.txt

Data Checking

Temporal Data Check : Ignore cases and controls that are outside the specified temporal study period.

Temporal Window

Maximum Temporal Cluster Size : 50 percent of study period

Space And Time Adjustments

Temporal Adjustment : None
Adjust for known relative risks : No

Inference

P-Value Reporting : Default Combination
Adjusting for More Likely Clusters : No
Number of Replications : 999

Additional Output

Report Critical Values : No
Report Monte Carlo Rank : No

Run Options

Processor Usage : All Available Proccessors
Logging Analysis : Yes
Suppress Warnings : No

Program completed : Sat Jan 5 15:29:52 2013
Total Running Time : 11 seconds
Processor Usage : 2 processors

A.4 Ano de 2010

SaTScan v9.1.1

Program run on: Sat Jan 5 15:30:22 2013

Retrospective Purely Temporal analysis
scanning for clusters with high rates
using the Discrete Poisson model.

SUMMARY OF DATA

Study period.....: 2010/1/1 to 2010/12/31
Total population.....: 190811860
Total number of cases.....: 984689
Annual cases / 100000.....: 516.4

MOST LIKELY CLUSTER

1.Location IDs included.: All
Time frame.....: 2010/1/1 to 2010/5/31
Number of cases.....: 794440
Expected cases.....: 407222.08
Annual cases / 100000.: 1007.4
Observed / expected...: 1.95
Relative risk.....: 5.92
Log likelihood ratio...: 319670.620340
P-value.....: 0.001

PARAMETER SETTINGS

Input

Case File : /Users/rafaellewkowicz/Downloads/dengue/dengue.cas
Population File : /Users/rafaellewkowicz/Downloads/dengue/dengue.pop
Time Precision : Month
Start Time : 2010/1/1
End Time : 2010/12/31

Analysis

Type of Analysis : Retrospective Purely Temporal
Probability Model : Discrete Poisson
Scan for Areas with : High Rates
Time Aggregation Units : Month
Time Aggregation Length : 1

Output

Results File : /Users/rafaellewkowicz/Downloads/dengue/resultados/Resultados 2010.txt

Data Checking

Temporal Data Check : Ignore cases and controls that are outside the specified temporal study period.

Temporal Window

Maximum Temporal Cluster Size : 50 percent of study period

Space And Time Adjustments

Temporal Adjustment : None
Adjust for known relative risks : No

Inference

P-Value Reporting : Default Combination
Adjusting for More Likely Clusters : No
Number of Replications : 999

Additional Output

Report Critical Values : No
Report Monte Carlo Rank : No

Run Options

Processor Usage : All Available Processors
Logging Analysis : Yes
Suppress Warnings : No

Program completed : Sat Jan 5 15:30:37 2013
Total Running Time : 15 seconds
Processor Usage : 2 processors

A.5 Ano de 2011

SaTScan v9.1.1

Program run on: Sat Jan 5 15:31:02 2013

Retrospective Purely Temporal analysis
scanning for clusters with high rates
using the Discrete Poisson model.

SUMMARY OF DATA

Study period.....: 2011/1/1 to 2011/12/31
Total population.....: 192434037
Total number of cases.....: 711024
Annual cases / 100000.....: 369.7

MOST LIKELY CLUSTER

1.Location IDs included.: All
Time frame.....: 2011/2/1 to 2011/6/30
Number of cases.....: 605684
Expected cases.....: 292101.95
Annual cases / 100000.: 766.7
Observed / expected...: 2.07
Relative risk.....: 8.25
Log likelihood ratio...: 296277.399621
P-value.....: 0.001

PARAMETER SETTINGS

Input

Case File : /Users/rafaellewkowicz/Downloads/dengue/dengue.cas
Population File : /Users/rafaellewkowicz/Downloads/dengue/dengue.pop
Time Precision : Month
Start Time : 2011/1/1
End Time : 2011/12/31

Analysis

Type of Analysis : Retrospective Purely Temporal
Probability Model : Discrete Poisson
Scan for Areas with : High Rates
Time Aggregation Units : Month
Time Aggregation Length : 1

Output

Results File : /Users/rafaellewkowicz/Downloads/dengue/resultados/Resultados 2011.txt

Data Checking

Temporal Data Check : Ignore cases and controls that are outside the specified temporal study period.

Temporal Window

Maximum Temporal Cluster Size : 50 percent of study period

Space And Time Adjustments

Temporal Adjustment : None
Adjust for known relative risks : No

Inference

P-Value Reporting : Default Combination
Adjusting for More Likely Clusters : No
Number of Replications : 999

Additional Output

Report Critical Values : No
Report Monte Carlo Rank : No

Run Options

Processor Usage : All Available Processors
Logging Analysis : Yes

Suppress Warnings : No

Program completed : Sat Jan 5 15:31:15 2013
Total Running Time : 13 seconds
Processor Usage : 2 processors

A.6 Intervalo completo 2007 - 2011 (Scan Temporal)

SaTScan v9.1.1

Program run on: Sat Jan 5 15:16:31 2013

Retrospective Purely Temporal analysis
scanning for clusters with high rates
using the Discrete Poisson model.

SUMMARY OF DATA

Study period.....: 2007/1/1 to 2011/12/31
Total population.....: 190742664
Total number of cases.....: 3142080
Annual cases / 100000.....: 329.5

MOST LIKELY CLUSTER

1. Location IDs included.: All
Time frame.....: 2010/1/1 to 2010/5/31
Number of cases.....: 794440
Expected cases.....: 259835.85
Annual cases / 100000.: 1007.4
Observed / expected...: 3.06
Relative risk.....: 3.75
Log likelihood ratio...: 406217.477211
P-value.....: 0.001

PARAMETER SETTINGS

Input

Case File : /Users/rafaellewkowicz/Downloads/dengue/dengue.cas

Population File : /Users/rafaellewkowicz/Downloads/dengue/dengue.pop
Time Precision : Month
Start Time : 2007/1/1
End Time : 2011/12/31

Analysis

Type of Analysis : Retrospective Purely Temporal
Probability Model : Discrete Poisson
Scan for Areas with : High Rates
Time Aggregation Units : Month
Time Aggregation Length : 1

Output

Results File : /Users/rafaellewkowicz/Downloads/dengue/resultados/Resultados.txt

Data Checking

Temporal Data Check : Ignore cases and controls that are outside the specified temporal study period.

Temporal Window

Maximum Temporal Cluster Size : 50 percent of study period

Space And Time Adjustments

Temporal Adjustment : None
Adjust for known relative risks : No

Inference

P-Value Reporting : Default Combination
Adjusting for More Likely Clusters : No
Number of Replications : 999

Additional Output

Report Critical Values : No
Report Monte Carlo Rank : No

Run Options

Processor Usage : All Available Processors
Logging Analysis : Yes
Suppress Warnings : No

Program completed : Sat Jan 5 15:17:01 2013
Total Running Time : 30 seconds
Processor Usage : 2 processors

A.7 Intervalo completo 2007 - 2011 (Scan Espacial)

SaTScan v9.1.1

Program run on: Fri Jan 18 22:43:07 2013

Retrospective Space-Time analysis
scanning for clusters with high rates
using the Discrete Poisson model.
Analysis includes purely spatial and purely temporal clusters.

SUMMARY OF DATA

Study period.....: 2007/1/1 to 2011/12/31
Number of locations.....: 5563
Total population.....: 190742862
Total number of cases.....: 3142080
Annual cases / 100000.....: 329.5

MOST LIKELY CLUSTER

1.Location IDs included.: 510625, 510260, 510020, 510628,
510617, 510390, 510270, 510100,
510779, 510180, 520170, 510665,
520250, 520380, 510310, 510820,
520340, 520310, 521370, 521925,
510630, 510385, 520215, 521400,
521295, 510718, 510810, 521720,
510719, 521220, 520235, 520710,
521200, 521483, 521100, 520725,
510704, 520753, 510706, 521520,
521020, 520760, 510670, 510420,
521030, 521405, 521565, 510700,
510120, 520090, 521385, 520640,
510645, 520440, 520570, 522157,
520393, 521470, 521340, 520929,
521160, 520890, 521900, 521120,

521890, 521390, 510620, 510040,
522020, 522028, 510729, 520945,
521970, 520420, 510185, 520495,
521015, 520357, 520500, 520960,
522010, 521690, 521810, 520260,
510360, 522170, 521090, 521486,
510631, 520540, 510788, 520015,
520780, 521860, 510030, 521940,
520085, 510520, 510740, 520980,
521040, 510637, 510480, 510760,
521060, 510267, 522005, 521870,
521645, 522150, 521487, 521310,
521935, 521640, 520130, 521056,
522015, 510776, 520055, 522100,
521375, 521590, 170200, 521440,
510460, 520082, 521140, 520160,
520280, 521950, 521525, 522160,
521180, 520470, 521410, 510035,
521205, 521280, 521945, 521570,
171884, 521170, 520860, 521910,
521000, 521990, 521680, 510624,
521190, 520520, 520750, 510060,
520680, 520460, 521971, 520013,
520320, 521800, 520360, 520995,
510300, 520880, 521960, 510370,
520545, 521500, 522140, 521540,
521880, 522230, 521973, 510785,
520920, 521450, 521730, 520005,
520810, 522190, 520740, 520485,
521930, 520180, 520870, 522119,
520551, 520580, 510269, 522050,
520110, 172097, 520140, 522145,
510340, 510830, 520840, 500793,
510452, 510850, 510622, 520547,
522155, 522045, 520465, 520735,
521460, 510840, 520970, 521805,
510780, 510160, 521377, 520355,
521260, 510525, 520145, 510530,
520010, 520455, 521300, 170070,
510590, 510677, 500025, 520650,
510735, 510792, 510010, 510724,
522205, 510770, 500325, 510350,
521839, 521560, 520030, 170820,
520505, 521230, 521770, 510305,
510050, 170765, 510610, 520330,
520815, 500640, 510490, 520430,
521305, 520025, 521210, 510335,
520050, 522060, 500295, 521710,

510790, 510558, 521975, 171575,
170386, 521308, 171150, 520552,
500390, 521850, 521130, 522200,
522026, 172085, 520410, 510600,
520150, 520630, 520915, 510130,
520350, 170950, 521380, 521523,
510730, 520910, 510650, 520017,
520993, 522185, 172025, 530010,
521080, 500330, 510885, 510800,
500290, 170730, 510726, 520549,
510774, 521250, 510345, 521630,
521600, 171660, 521530, 510777,
521920, 521225, 510455, 521580,
521878, 313140, 521760, 522040,
510170, 522000, 315980, 520390,
510619, 171190, 520425, 520060,
310980, 510685, 521740, 520450,
170035, 510623, 521150, 171620,
510860, 311260, 510454, 520530,
172049, 310375, 520020, 500740,
522180, 521290, 510805, 510890,
170625, 510320, 520800, 311580,
510795, 522108, 520590, 520620,
311180, 521010, 500440, 171889,
510642, 313420, 170980, 170610,
500260, 500630, 522220, 520790,
510560, 313862, 312910, 170755,
500769, 521480, 510621, 171550,
170370, 317043, 171890, 171750,
310945, 316960, 311455, 510880,
171500, 520480, 510250, 510410,
520400, 314280, 520850, 521490,
172065, 500730, 171420, 520660,
520120, 170510, 171845, 510626,
510343, 510263, 313440, 170460,
510523, 520510, 510794, 521350,
170560, 171820, 317040, 170240,
171250, 500100, 352965, 354040,
520990, 354570, 500150, 311110,
510720, 510775, 354610, 521550,
510562, 171610, 310350, 353475,
354450, 520080, 354740, 352070,
520490, 355530, 500310, 315280,
353590, 522068, 150670, 510710,
171360, 350395, 171370, 354660,
353000, 351420, 355490, 522130,
351800, 520396, 520940, 316130,
170710, 500780, 317020, 520690,

354720, 500490, 172093, 510279,
354765, 500020, 311500, 310930,
355695, 355580, 500750, 510680,
510125, 510627, 510715, 521830,
352044, 510395, 171700, 352820,
314700, 352480, 312790, 353284,
171515, 354420, 351070, 170390,
354900, 312350, 522070, 351520,
170040, 312620, 311690, 312860,
500710, 170025, 170555, 314437,
353520, 350260, 520830, 352910,
313340, 351550, 172100, 355255,
353690, 520670, 171215, 170310,
521980, 510682, 171800, 354920,
510025, 352300, 354025, 351385,
510190, 317047, 521270, 312480,
352960, 510450, 353660, 353625,
354030, 510510, 353740, 313070,
310010, 170270, 350120, 171790,
354925, 355610, 500830, 310450,
350180, 500270, 355230, 314310,
355710, 312247, 172090, 170700,
351890, 317100, 510380, 353420,
315640, 311930, 171900, 500800,
311140, 171865, 351100, 350420,
171525, 312710, 351690, 171780,
171200, 170720, 315070, 350210,
351290, 352830, 315445, 317110,
510500, 171330, 150658, 314500,
510787, 351590, 353500, 170110,
313160, 313710, 312700, 353286,
171195, 353210, 171320, 510629,
353260, 355130, 310820, 172110,
351980, 313753, 351780, 351680,
170190, 311730, 353320, 315160,
355340, 353330, 510835, 171870,
354805, 351210, 310855, 311615,
314920, 353300, 352810, 353100,
315770, 317010, 317052, 353040,
171510, 353010, 353990, 500230,
353140, 352725, 352650, 353400,
350480, 500520, 352115, 354930,
500348, 355520, 350090, 355570,
500755, 353640, 500110, 500320,
353160, 500070, 351790, 314810,
313630, 353310, 353250, 354710,
355630, 353540, 353270, 353030,
314980, 500790, 350810, 310070,

355715, 170825, 352450, 510268,
355510, 350620, 312890, 510675,
500560, 354980, 353960, 351820,
170830, 350280, 354440, 315340,
350775, 351750, 350460, 314800,
351740, 353480, 510895, 351440,
352570, 171875, 312125, 352970,
291955, 351130, 350650, 352600,
353390, 353280, 171670, 350550,
312070, 171665, 352010, 170930,
510336, 350300, 355535, 352160,
350640, 353490, 351250, 350510,
355560, 354080, 315760, 172125,
311820, 510618, 315690, 316680,
355190, 351710, 171650, 150613,
313750, 352950, 170330, 510515,
150276, 351600, 355260, 170600,
510285, 500190, 350020, 351200,
500600, 351940, 351580, 350930,
351650, 171110, 350820, 150270,
510080, 315057, 353730, 352420,
354510, 350770, 352130, 354130,
351120, 317075, 353325, 171270,
350010, 353770, 351190, 310400,
351495, 355600, 352410, 172015,
352740, 353150, 350910, 350440,
354360, 500215, 352150, 354840,
510615, 351492, 150555, 351770,
354323, 353570, 171840, 354480,
351512, 352890, 355440, 350110,
352080, 354150, 311430, 351110,
354380, 316420, 354830, 353830,
352540, 353510, 316170, 353600,
352770, 352885, 353460, 354160,
353700, 510330, 500580, 354085,
354460, 350610, 171180, 315213,
353810, 354940, 355680, 353190,
353900, 350335, 354470, 354770,
350080, 170360, 510550, 316110,
350370, 312950, 352190, 313005,
351920, 354310, 351720, 310380,
310825, 354120, 354560, 315550,
510517, 353430, 355690, 354180,
350580, 351320, 350890, 353350,
355365, 171888, 500720, 352710,
354950, 355310, 500540, 352870,
355320, 316810, 350130, 351560,
353950, 354490, 355500, 317000,

353360, 312960, 351700, 311640,
311115, 316890, 354140, 500080,
172130, 171240, 351620, 170410,
354270, 351010, 171090, 292890,
353130, 500220, 352920, 150125,
351900, 354020, 150304, 352060,
290810, 290320, 170320, 500085,
350880, 290930, 310940, 315120,
351730, 316210, 314120, 354240,
354010, 350740, 352430, 170550,
353020, 313868, 500620, 350560,
312970, 315300, 353920, 354200,
500410, 290740, 355390, 355370,
500200, 500350, 350240, 353630,
354220, 355170, 351810, 314225,
355590, 150616, 311783, 500500,
352270, 316935, 350590, 290910,
171630, 352510, 354000, 170230,
171050, 352370, 291735, 312120,
352560, 352580, 354170, 314270,
350780, 310700, 314915, 510337,
351860, 500470, 311150, 500450,
171850, 313520, 354650, 316265,
350470, 353220, 311560, 351460,
314350, 314640, 353410, 355290,
500345, 350720, 110030, 310860,
354250, 315970, 311240, 354550,
351400, 292620, 350140, 354340,
353890, 313730, 291077, 354090,
170305, 351960, 351530, 311510,
355430, 313865, 291110, 500797,
355270, 316660, 352900, 351535,
351910, 411260, 317080, 353450,
352790, 352930, 500370, 316510,
313535, 314130, 290140, 353205,
355660, 353550, 150840, 350100,
312470, 354110, 354425, 500400,
313695, 411090, 355150, 351885,
311880, 315370, 351670, 290250,
312965, 353290, 500380, 500840,
412360, 316430, 411830, 316820,
310020, 351310, 313210, 410710,
316245, 150808, 316225, 350430,
351470, 315290, 292905, 292810,
353215, 110003, 171488, 150543,
500510, 313560, 354370, 150730,
317060, 311980, 312320, 350340,
412450, 351685, 352200, 355140,

412730, 314200, 412420, 220300,
411130, 316470, 351660, 171570,
290610, 150034, 411030, 313375,
351990, 411500, 310510, 110006,
352880, 353370, 314795, 290970,
354690, 354790, 317090, 352780,
411380, 354625, 500625, 411710,
350170, 313810, 412000, 410340,
351565, 313930, 316240, 412020,
350320, 150775, 350150, 314790,
350520, 313290, 314085, 350940,
316220, 410590, 410510, 313880,
410080, 352760, 412490, 312630,
412590, 220290, 355090, 311650,
350400, 150503, 351330, 352750,
353970, 220130, 312570, 350670,
351450, 350980, 292820, 410800,
312660, 355395, 314050, 510140,
350730, 351090, 500090, 411640,
355550, 411810, 350600, 410890,
353715, 412050, 221062, 314320,
351000, 411350, 355475, 350830,
410670, 313480, 500240, 354760,
350680, 220975, 500525, 352290,
290940, 292020, 353940, 410920,
290710, 312340, 171886, 412330,
411600, 110005, 411360, 293090,
170307, 316294, 353530, 411970,
412370, 351430, 351930, 351610,
315200, 314360, 412340, 171395,
220440, 353657, 293030, 354290,
410280, 170130, 210050, 314240,
313030, 353670, 220920, 310190,
412033, 411840, 315150, 412830,
313505, 313720, 351950, 354750,
410810, 311910, 500210, 314330,
350070, 310760, 354320, 352530,
500660, 410090, 310410, 313640,
290440, 311280, 355050, 410060,
316390, 500515, 412395, 310740,
220660, 292840, 310920, 412650,
354890, 411190, 412100, 310480,
312830, 410220, 510325, 353050,
351519, 412640, 411630, 312380,
317103, 311270, 150215, 110007,
110146, 411650, 170210, 411690,
412130, 351370, 354540, 350745,
291733, 313530, 355330, 412670,

312870, 410115, 315050, 170900,
220885, 411000, 411340, 411100,
354070, 314650, 110092, 352800,
500570, 351410, 170388, 313690,
500280, 310730, 314510, 352980,
293015, 310420, 410210, 411965,
411590, 316060, 350530, 350790,
412040, 352680, 312735, 411800,
410110, 312090, 293345, 352000,
410240, 170384, 354640, 313830,
353470, 412460, 312810, 355360,
410360, 354630, 316040, 410780,
412390, 412840, 311440, 310390,
354970, 410640, 313110, 410370,
412690, 411410, 412270, 410910,
312825, 411270, 410980, 351015,
350200, 311995, 353930, 311710,
412240, 410270, 351080, 314410,
314250, 314690, 355470, 350055,
411370, 313680, 315140, 352360,
352380, 410560, 170300, 355720,
314970, 412555, 312670, 350360,
411740, 351925, 350870, 412260,
311760, 170770, 353380, 314520,
410725, 220760, 314300, 411520,
412530, 210280, 352090, 293075,
500690, 350630, 351270, 411180,
355080, 412625, 410150, 354620,
312610, 412310, 411155, 411750,
290390, 313970, 351390, 411240,
411660, 313510, 410990, 314505,
314890, 411040, 352860, 314580,
410190, 411720, 355010, 313050,
411480, 500124, 354700, 410010,
355640, 310200, 310950, 171720,
500060, 354105, 412610, 411420,
352670, 170215, 412180, 314960,
314545, 410730, 311890, 315320,
412410, 351140, 310990, 314740,
150563, 292340, 316180, 412600,
355460, 310430, 410140, 352110,
411700, 150553, 412862, 355040,
172208, 312733, 412190, 410790,
311130, 411300, 411160, 292225,
411210, 211200, 351170, 410380,
412320, 412430, 353880, 354390,
316460, 350030, 312230, 411470,
310320, 410550, 500460, 350060,

310840, 316310, 150277, 310960,
411290, 315180, 312020, 411110,
110009, 312240, 312640, 354910,
314655, 313350, 313100, 350750,
350330, 313020, 350310, 355120,
350040, 313200, 354670, 292045,
220552, 350450, 410600, 411721,
314710, 410350, 350230, 310530,
150295, 410320, 411550, 315850,
315220, 172210, 150745, 412680,
315450, 311040, 311160, 312760,
351240, 311200, 110018, 313925,
412720, 310160, 311420, 351220,
311547, 410470, 310710, 293000,
410900, 500315, 355420, 316720,
412667, 311100, 351540, 411280,
310080, 316690, 411490, 412880,
411729, 410750, 410660, 312160,
292370, 220320, 412210, 412810,
352140, 352350, 170650, 312520,
410752, 312100, 110147, 355730,
311120, 412110, 312010, 411310,
314290, 412470, 313380, 352690,
313570, 410610, 412070, 354810,
290660, 312600, 314100, 315360,
292640, 220870, 412535, 311030,
353870, 412790, 412580, 411890,
171855, 355300, 310260, 312410,
351515, 353610, 351518, 312720,
210950, 316695, 500430, 291320,
312430, 312550, 411880, 353070,
311250, 314070, 311400, 292105,
311660, 410770, 315830, 410170,
311870, 310850, 310500, 316120,
411885, 312990, 311810, 350690,
352180, 500375, 315880, 291170,
411575, 310140, 354515, 410330,
314110, 410347, 353080, 313900,
412290, 313665, 351230, 350380,
500795, 315330, 170100, 315920,
410050, 312780, 411170, 110148,
110145, 314720, 311190, 412660,
354400, 171380, 412500, 316940,
350500, 355380, 354580, 315170,
351280, 410775, 314930, 316590,
292450, 410685, 411510, 410250,
313490, 313370, 353580, 314560,
350160, 312030, 410430, 411920,

352260, 353750, 411230, 292160,
 293260, 312540, 410760, 170105,
 412780, 220110, 290475, 412620,
 291340, 313010, 411610, 354050,
 315460, 311787, 410070, 410970,
 351260, 352640, 311450, 315900,
 110004, 316292, 310670, 353090,
 351905, 311470, 411060, 311350,
 500480, 353340, 411375, 172000,
 316045, 210405, 410655, 314460,
 352280, 410755, 355465, 170030,
 313220, 314990, 412400, 500568,
 316710, 352585, 210140, 313150,
 411342, 354800, 315990, 316295,
 355240, 353650, 316020, 312360,
 314015, 411220, 313760, 316230,
 314600, 171430, 292320, 315530,
 316553, 317120, 410832, 314340,
 350050, 355450, 210407, 354210,
 410870, 311860, 412850, 351040,
 315060, 352470, 316580, 351160,
 310900, 315470, 311900, 351907,
 310240, 315560, 317070, 352700,
 316980, 291980, 410855, 411250,
 314345, 290410, 291940, 410700,
 410860, 354165, 313060, 316650,
 312060, 353180, 313460, 351850,
 351150, 312440, 171245, 312980,
 110029, 354350, 350220, 110037,
 110120, 311790, 410300, 355160,
 410880, 290520, 312707, 353710,
 410337, 314770, 412540, 310810,
 315780, 411150, 311390, 314260,
 313820, 313040, 313835, 291740,
 310830, 311750, 170382, 351490,
 411373, 316740, 350190, 310620,
 316970, 220190, 220323, 290420,
 310910, 314625, 310800, 412065,
 220665, 210725, 110028

Coordinates / radius...: (14.669200 S, 52.361200 W) / 1070.77 km
 Time frame.....: 2010/1/1 to 2010/5/31
 Population.....: 46119247
 Number of cases.....: 514089
 Expected cases.....: 62924.71
 Annual cases / 100000.: 2691.9
 Observed / expected...: 8.17
 Relative risk.....: 9.57
 Log likelihood ratio...: 663454.690137

P-value.....: < 0.000000000000000010

SECONDARY CLUSTERS

2.Location IDs included.: All

Time frame.....: 2010/1/1 to 2010/5/31

Number of cases.....: 794440

Expected cases.....: 259835.77

Annual cases / 100000.: 1007.4

Observed / expected...: 3.06

Relative risk.....: 3.75

Log likelihood ratio...: 406217.655403

P-value.....: < 0.000000000000000010

PARAMETER SETTINGS

Input

Case File : C:\Users\Renato\Desktop\Rafael Dengue\dengue.cas
Population File : C:\Users\Renato\Desktop\Rafael Dengue\dengue.pop
Coordinates File : C:\Users\Renato\Desktop\Rafael Dengue\BR.geo
Time Precision : Month
Start Time : 2007/1/1
End Time : 2011/12/31
Coordinates : Latitude/Longitude

Analysis

Type of Analysis : Retrospective Space-Time
Probability Model : Discrete Poisson
Scan for Areas with : High Rates
Time Aggregation Units : Month
Time Aggregation Length : 1

Output

Results File : C:\Users\Renato\Desktop\Rafael Dengue\Resultados Espacial\Spatial2007-2011.txt

Data Checking

Temporal Data Check : Ignore cases and controls that are outside the specified temporal study period.
Geographical Data Check : Ignore observations that are outside the specified geographical area.

Spatial Neighbors

Use Non-Euclidian Neighbors file : No
Use Meta Locations File : No
Multiple Coordinates Type : Allow only set of coordinates per location ID.

Spatial Window

Maximum Spatial Cluster Size : 50 percent of population at risk
Include Purely Temporal Clusters : Yes
Window Shape : Circular

Temporal Window

Maximum Temporal Cluster Size : 50 percent of study period
Include Purely Spatial Clusters : Yes

Space And Time Adjustments

Temporal Adjustment : None
Spatial Adjustment : None
Adjust for known relative risks : No

Inference

P-Value Reporting : Default Combination
Adjusting for More Likely Clusters : No
Number of Replications : 999

Clusters Reported

Criteria for Reporting Secondary Clusters : No Geographical Overlap

Additional Output

Report Critical Values : No
Report Monte Carlo Rank : No

Run Options

Processor Usage : All Available Processors
Logging Analysis : Yes
Suppress Warnings : No

Program completed : Sat Jan 19 15:58:35 2013
Total Running Time : 17 hours 15 minutes 28 seconds
Processor Usage : 6 processors

Referências Bibliográficas

- Emma K. Chaput. Spatial analysis of human granulocytic ehrlichiosis near lyme, connecticut. *Emerging Infectious Diseases*, 8:943–948, 2002.
- DataSUS. Departamento de informática do sus, Junho 2012. URL <http://www2.datasus.gov.br/>.
- George S. Fishman. *Monte Carlo: Concepts, Algorithms, and Applications*. New York: Springer, 1995.
- Jayro Guimarães Jr. *Biossegurança e controle de infecção cruzada*. Santos, Livraria Editora, 2001.
- Joseph Hair. *Multivariate Data Analysis, Fifth edition*. Prentice Hall, Inc., 1998.
- Martin Kulldorff. A spatial scan statistic. *Communications in Statistics: Theory and Methods*, 26:1481–1496, 1997.
- Martin Kulldorff. Prospective time-periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society*, A164:61–72, 2001.
- Nicholas Metropolis. The monte carlo method. *Journal of the American Statistical Association*, 44:335–341, 1949.
- Kulldorff Nagarwalla. Spatial disease clusters: Detection and inference. *Statistics in Medicine*, 14:799–810, 1995.
- Frank B. Osei. Spatial dependency of v. cholera prevalence on open space refuse dumps in kumasi. ghana: a spatial statistical modeling. *International Journal of Health Geographics*, page 7:62, 2008.
- Débora Cruz Payão Pellegrini. Análise espaço-temporal da leptospirose no município do rio de janeiro (1995-1999). *Rio de Janeiro: Fundação Oswaldo Cruz*, 2002.
- R. Software for general statistics, Julho 2012. URL www.r-project.org/.
- SatScan. Software for the spatial, temporal, and space-time statistics, Junho 2012. URL <http://www.satscan.org/>.